Cognition 190 (2019) 199–211

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Evidence of stable individual differences in implicit learning

Priya B. Kalra[a,*], John D.E. Gabrieli[b], Amy S. Finn[c]

[a] Waisman Center, University of Wisconsin—Madison, United States
[b] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States
[c] Department of Psychology, University of Toronto, Canada

ARTICLE INFO

Keywords:
Implicit learning
Procedural memory
Reliability
Individual differences
Memory systems

ABSTRACT

There is a fundamental psychological and neuropsychological distinction between explicit and implicit memory, and it has been proposed that whereas there are stable trait individual differences in explicit memory ability, there are not such differences across people for implicit learning. There is, however, little evidence about whether or not there are stable trait differences in implicit learning. Here we performed a test-retest reliability study with healthy young adults in which they performed four implicit learning tasks (artificial grammar learning, probabilistic classification, serial response, and implicit category learning) twice, about a week apart. We found medium (by Cohen's guidelines) test-retest reliability for three of the tasks: probabilistic classification, serial response, and implicit category learning, suggesting that differences in implicit learning ability are more stable than originally thought. In addition, implicit learning on all tasks was unrelated to explicit measures: we did not find any correlation between implicit learning measures and independent measures of IQ, working memory, or explicit learning ability. These findings indicate that implicit learning, like explicit learning, varies reliably across individuals.

1. Introduction

Implicit learning was originally defined as unconscious learning that yields abstract knowledge (Reber, 1989). It was further hypothesized that implicit learning would have five characteristics: (1) age independence, (2) robustness to disease or injury, (3) low variability across individuals, (4) IQ independence, and (5) conservation across phylogeny (Reber, 1993). Considerable evidence supports the notion that implicit learning can be robust to disease or injury (Reber, 2013), and that it is relatively age-independent, maturing earlier in development than explicit memory (Finn et al., 2015) and diminishing less with aging (Churchill, Stanis, Press, Kushelev, & Greenough, 2003; Drag & Bieliauskas, 2010; Vakil & Agmon-Ashkenazi, 1997). However, there is little empirical evidence that supports or contradicts the assertion of little or no stable individual differences in implicit learning. A direct measure of individual stability for a particular form of implicit learning would involve a test/re-test design to measure whether or not there is individual stability in such learning over time. Further, if there are stable individual differences in implicit learning on a given task, it is unknown if there is a broader implicit learning ability that would generalize across different measures of implicit learning and different modalities, cognitive and perceptual-motor. The present study was designed to address this gap of knowledge about stable individual differences in implicit learning. This gap in our understanding is striking given that implicit learning is recognized as a core system that underlies learning in multiple domains including language (Ullman, 2001), music (Creel, Newport, & Aslin, 2004), and even learning about the statistical structure of our environments (Batterink, Reber, Neville, & Paller, 2015).

The idea that that implicit learning can be cognitively and neurobiologically distinct from explicit forms of learning and memory has been supported by research showing dissociations via neuropsychology, neuroimaging, and behavioral experiments. In particular, neuropsychological studies of patients with brain lesions have documented that multiple forms of implicit learning are preserved in amnesia (Foerde & Shohamy, 1998; Graf & Schacter, 1985; Knowlton, Ramus, & Squire, 1992; Knowlton, Squire, & Gluck, 1994; Meulemans & der Linden, 2000, 2003) or Alzheimer's disease (Gabrieli, Corkin, Mickel, & Growdon, 1993; Nosofsky, Denton, Zaki, Murphy-Knudsen, & Unverzagt, 2012; Reber, Martinez, & Weintraub, 2003) despite the severe impairment of explicit memory in these patients. Thus, these forms of implicit learning can be accomplished without the hippocampus and associated medial temporal-lobe structures that are injured in these memory disorders. Conversely, these forms of implicit learning are often impaired in disorders of the basal ganglia and cerebellum (Ashby, Valentin, Johnson, & Stout, 2005; Foerde & Shohamy, 2011a,2011b;

* Corresponding author at: Waisman Center, University of Wisconsin—Madison, 1500 Highland Avenue, Madison, WI 53705, United States.
E-mail address: pkalra2@wisc.edu (P.B. Kalra).

https://doi.org/10.1016/j.cognition.2019.05.007
Received 19 July 2018; Received in revised form 6 May 2019; Accepted 7 May 2019
Available online 16 May 2019
0010-0277/

Knowlton, Mangels, & Squire, 1996; Knowlton et al., 1996; Shohamy, Myers, Onlaor, & Gluck, 2004; Smith & McDowall, 2006), which suggests that these neural systems support such implicit learning. The precise relation between implicit and explicit learning in regards to neural systems is complex, however, behavior because patients with basal ganglia dysfunction can also be impaired on some measures of explicit learning (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Ell, Weinstein, & Ivry, 2010; Maddox, Aparicio, Marchant, & Ivry, 2005; Voytek & Knight, 2010), and, conversely, patients with medial temporal lobe dysfunction can be impaired on some measures of implicit learning (Knowlton et al., 1992, 1993; Warren & Duff, 2014; Zaki, 2004).

Nevertheless, there is ample evidence for considering these cognitive and neurobiological systems as largely distinct (Gabrieli, 1998). Neuroimaging studies have indicated that these forms of learning have distinct neural correlates (Poldrack & Foerde, 2008; Poldrack et al., 2001, 2005; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Seger, Prabhakaran, Poldrack, & Gabrieli, 2000; Willingham & Goedert-Eschmann, 1999; Willingham, Salidis, & Gabrieli, 2002). Behavioral measures (Cohen & Poldrack, 1997; Foerde, Poldrack, & Knowlton, 2007; Willingham et al., 1993) and manipulations—such as the use of a secondary task or modulation in the timing of feedback— have also distinguished implicit from explicit learning (Foerde & Shohamy, 2011a,2011b; Foerde et al., 2007; Heindel, Festa, Ott, Landy, & Salmon, 2013). Thus, converging lines of evidence support the understanding of implicit learning as a construct that can be dissociable from explicit or declarative forms of learning.

Further, while explicit learning changes greatly with aging and development (Ghetti, Angelini, & Annunzio, 2008), implicit learning changes less (if at all). Studies of normal aging suggest that while some decline in implicit learning may be present in older adults, the degree of impairment is much smaller than what is observed for explicit learning (Bhakuni & Mutha, 2015; Brown, Robertson, & Press, 2009; Curran, 1997; Fernandez-Ruiz, Hall, Vergara, & Diaz, 2000; Howard & Howard, 2013; Kuerten, De Vries, Kowal, Zwitserlood, & Floeel, 2012; Midford & Kirsner, 2005; Ofen et al., 2007; Rieckmann & Backman, 2009; Schugens, Daum, Spindler, & Birbaumer, 1997). Developmental dissociations between implicit and explicit learning have also been observed (Amso & Davidow, 2012; Janacsek, Fiser, & Nemeth, 2012; Meulemans, Van der Linden, & Perruchet, 1998; Thomas & Nelson, 2001; Thomas et al., 2004). One study used multiple measures of implicit learning and found that children at age 10 years demonstrated adult-like levels of implicit learning but lower levels of explicit learning (Finn et al., 2015). Additionally, developmental disabilities that affect general intelligence (and therefore affect explicit learning) may not impair implicit learning (Atwell, Conners, & Merrill, 2003; Bussy, Charrin, Brun, Curie, & des Portes, 2011; Vinter & Detable, 2008). Some research suggests that the developmental invariance of implicit learning could be an asset for child learners, allowing them to surpass adults in situations where explicit learning mechanisms are less effective than or could even interfere with implicit learning, such as some aspects of language learning (Finn, Lee, Kraus, & Hudson Kam, 2014; Janacsek et al., 2012; Ramscar & Gitcho, 2007; Ullman, 2001).

The relatively early development of implicit learning does not, however, mean that implicit learning functions optimally in all individuals all of the time. Deficits in implicit learning have been suggested to cause an array of learning disorders such as dyslexia, specific language impairment (SLI), and ADHD. The evidence to support these relationships between implicit learning and atypical development is mixed (Barnes, Howard, Howard, Kenealy, & Vaidya, 2010; Hedenius et al., 2011; Laasonen et al., 2014; Lukács & Kemény, 2014; Lum, Ullman, & Conti-Ramsden, 2013; Menghini et al., 2008; Pavlidou, Williams, & Kelly, 2009; Rosas et al., 2010; Rüsseler, Gerth, & Münte, 2006; Staels & Van den Broeck, 2017; Vloet et al., 2010), but these findings raise the possibility of a developmental basis for stable individual differences in implicit learning. Indeed, the underlying assumption of this work is that implicit learning is a stable trait that varies across individuals, with considerable implications for how children read, learn, and attend.

Despite its importance, research on individual differences in implicit learning is minimal and findings are mixed. In terms of cross-task correlations, two independent studies examined multiple measures of implicit learning and found that they did not correlate within an individual (Gebauer & Mackintosh, 2007; Horan et al., 2008). A few studies have tried to examine individual differences in implicit learning using the "criterion validation method," in which the correlation between the trait of unknown stability and a trait of known stability (in this case, IQ) is calculated (Fletcher, Maybery, & Bennett, 2000; Kaufman et al., 2010; Lin, 2004; Reber, Walkenfeld, & Hernstadt, 1991). If the trait in question is correlated with the stable trait, it can be assumed to be stable. However, if it is not correlated, then it cannot be assumed to be unstable because there are multiple reasons why something might not correlate with the stable trait chosen as the "criterion." Given this, the most widely accepted and statistically sound method for establishing whether there are stable individual differences in a trait is to establish the test-retest reliability of the measure of that trait (Cattell, 1971; Cronbach & Meehl, 1955; Cronbach, 1970).

Along these lines, one study of implicit motor learning has examined test-retest reliability using three tasks—a visuomotor adaptation task, an alternating serial reaction time task, and a non-alternating serial reaction time task (Stark-Inbar, Raza, Taylor, & Ivry, 2017). There were reliable individual differences across test sessions for visuomotor adaptation learning and for the alternating serial reaction time task, and reliable individual differences across test sessions for learning in the non-alternating serial reaction task when learning was measured mid-training, but not when measured at the end of training (possibly because fatigue influenced performance differently in participants across testing times). Overall, these findings suggest that perceptual-motor learning ability may be a stable difference across individuals, but does not speak to implicit learning more broadly, including non-perceptual-motor measures. Moreover, it is unknown as to how the test-retest reliability of implicit learning compares to explicit test-retest reliability.

In the present study, we used a test-retest design to test whether there are stable trait individual differences in implicit learning as measured by four tasks—artificial grammar learning, probabilistic classification, serial reaction time and category learning—in which implicit learning has been dissociated from explicit memory (Knowlton & Squire, 1993; Knowlton et al., 1992, 1994; Meulemans & der Linden, 2000, 2003; Nissen & Bullemer, 1987; Zannino et al., 2012) and which span perceptual-motor and cognitive domains. Participants performed all four tasks twice, 1–2 weeks apart. All participants also performed an explicit learning task at both visits (California Verbal Learning Test). In addition, participants were characterized on measures known to yield stable individual differences, namely IQ and working memory. Given the previously established stability and reliability of explicit learning measures (Rönnlund, Nyberg, Bäckman, & Nilsson, 2005; Waters & Caplan, 2003; Xu, Adam, Fang, & Vogel, 2018), as well as the fact that the California Verbal Learning Test is a standardized psychometric instrument, we expected to find high test-retest reliability for the California Verbal Learning Test. What was unknown was (1) whether or not people would show trait-like stability in implicit learning on each task across test and re-test sessions, and (2) whether implicit learning ability would be correlated across tasks, which would suggest a shared mechanism supporting such learning across tasks; and (3) whether implicit learning is dissociable from explicit memory and general intelligence (IQ) abilities.

2. Methods

2.1. Participants

Given an alpha level of 0.05, to reach 80% power assuming an effect

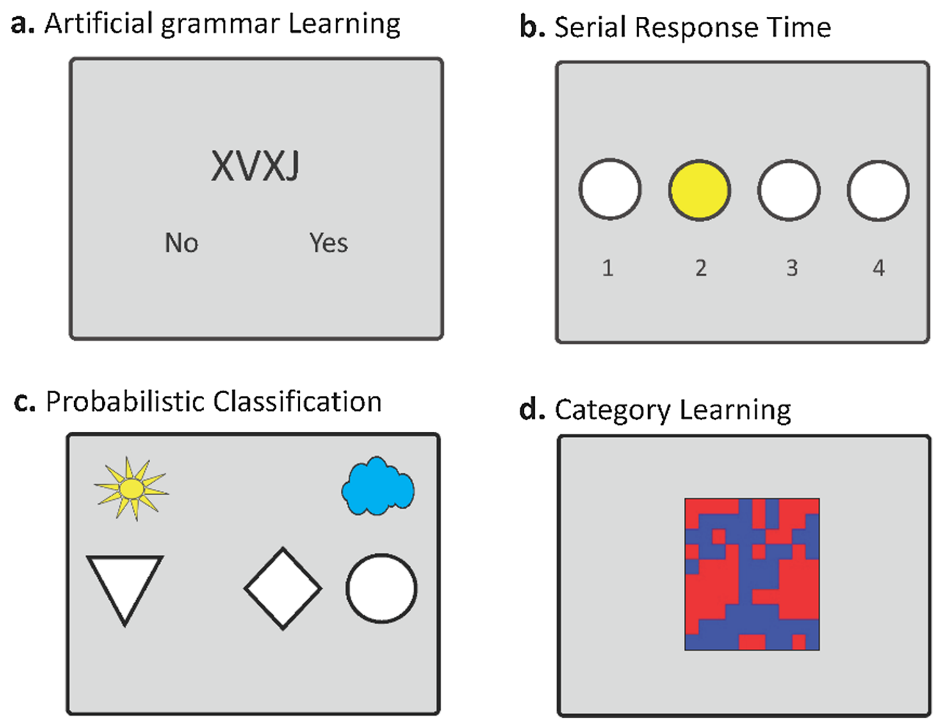

Fig. 1. Depictions of each of the four implicit learning tasks: (a) artificial grammar learning,(b) serial reaction time task,(c) probabilistic classification task, and (d) implicit category learning.

size of at least 0.3, our needed sample size was 64. However, since we anticipated attrition due to multiple test days, we committed to running 20% more, stopping at 76 participants. In total, 76 healthy young adults (mean age = 17.69, range = 16–22, 38 female) participated. Of these, 68 completed all tasks at both of two testing sessions. Reasons for not completing all tasks at both sessions were: did not return for second session ($n = 5$), program crashed ($n = 2$), and experimenter error ($n = 1$). Participants received course credit for participating.

2.2. Procedure

At each testing session, participants completed one of two versions of each of the following tasks in the following order: artificial grammar learning, serial response, probabilistic category learning, implicit category learning, and the California Verbal Learning Test. Versions were counterbalanced across participants. During the second visit only, participants also completed the Kaufman Brief Intelligence Test and a working memory span measure after completing the five repeated tasks listed above. All tests are detailed fully below. Sessions were separated by a minimum of one week and no more than one month (mean time between sessions = 13.89 days, *SD* = 2.39 days).

2.3. Materials and measures

2.3.1. Implicit learning measures

2.3.1.1. Artificial grammar learning. This task was based on artificial grammar learning tasks originally used by Reber (1967). Two independent versions (A and B) were created using two different Markov chains to produce letter strings. For both versions, there was an initial study phase in which participants viewed a series of letter strings on a computer screen and were instructed to write them down. There was no time limit. After copying each string, participants were told to cover their response before moving on to the next string. An experimenter was present in the room to ensure compliance on each trial with this covering procedure. A total of 23 study strings were generated and presented twice (each of the 23 strings were presented in random order once and then cycled through randomly again for a total of 46 training trials). Immediately after training, the copying paper and covering sheet were removed and participants were asked—unexpectedly—to decide whether new strings were grammatical. Of these new strings, 16 were grammatical and 16 were not, and half of each kind were high-chunk strength and half were low-chunk strength (high chunk strength = frequent letter pairings during the study phase see e.g., Pothos, 2007). We sampled evenly from the distribution of possible strings, including 16 short strings (2–4 items) and 16 longer strings (5–6 items) which were the same for all participants, and including a variety of paths through the Markov chains. All training stimuli (for each version) are included in Supplemental Fig. 1. The Markov chains used to create the stimuli can also be found in Supplemental Fig. 1. The dependent measure was the sensitivity index (“d prime”, Macmillan & Creelman, 2005) derived from the standardized hit rate (correct endorsement rate of grammatical test items) and false alarm rate (incorrect endorsement rate of non-grammatical test items).

2.3.1.2. Probabilistic classification task. This task was modeled after prior weather prediction tasks (Knowlton et al., 1994). Participants viewed a series of 15 unique cue combinations (which were comprised of between one and four simultaneously presented images) on a computer screen and were asked which of two “outcomes” the cue combination predicted. (See the Supplemental Table 1) for exact specifications of combinations, their frequencies and outcome probabilities. In version A, the cover story was like the traditional weather prediction task, the cues were different shapes of the same color, and the outcomes were “rain” or “sun.” In version B, the cover story was about baseball; the cues were pentagons of different colors, and the outcomes were “Team 1 wins” or “Team 2 wins.” In both versions, each cue always appeared in the same location. At first, participants did not know which outcome was predicted by any given cue combination; after deciding, they were given feedback in the form of a smiling or frowning face. Participants completed 252 trials in each version (this number was chosen to permit adequate item and item-feedback frequencies), broken into four roughly even blocks (63 trials each); participants were allowed brief breaks between blocks. The dependent measure was the percentage of chosen optimal outcomes (which outcome was more probable for each cue combination given prior feedback). See Fig. 1 for an illustration.

2.3.1.3. Serial reaction time task. This task was modeled after the serial reaction time tasks developed by Nissen and Bullemer (1987). On each trial, participants viewed a circle on a screen that could appear in one of 4 locations. They were instructed to press a key that corresponded to the location of each circle as quickly as possible. At each session, participants completed eight blocks of 96 trials each; in each version, blocks 1, 3, 7, and 8 were "random" blocks meaning that the circle appeared in a random sequence of locations with the caveat that locations did not repeat. Blocks 2, 4, 5 and 6 consisted of the same repeating 12-item sequence. For each version, a different second-order conditional 12-item sequence was used. The sequence for version A (121342314324) was taken from Reed and Johnson (1994). The sequence for version B (134213214243) was constructed following guidelines from Willingham and Dumas (1996). No feedback was provided to participants and a constant 250 ms response-stimulus interval followed each response. The dependent measure was a skill score. The skill score for each participant was calculated based on the mean RT for the final two sequence blocks (5,6) and the mean RT for the sequence-flanking random blocks (3 and 7) as follows: $\frac{MeanRT_{RandomBlocks(3,7)} - MeanRT_{SequenceBlocks(5,6)}}{MeanRT_{RandomBlocks(3,7)}}$ (see Galea, Albert, Ditye, & Miall, 2010).

We also assessed participants' explicit sequence knowledge from the serial response task. After their Session 2 serial reaction time task only (no explicit knowledge post-test was performed during Session 1 in order to avoid developing explicit strategies during the second session), participants were asked if they observed any kind of pattern in the task stimuli, and only if so, they were asked to produce (by free recall) as many items as possible using either numerical position indicators (1,2,3,4) or the keyboard keys that had been used during the task (f,g,h,j). To score an item as correctly recalled, it had to occur within a series of 3 or more correctly sequenced items (Willingham & Dumas, 1996). For this reason, scores can be 0 (no items recalled) or range from 3 to 12 items (i.e. producing the entire sequence).

2.3.1.4. Implicit category learning. This task is based on the prototype distortion category learning task first used by Posner and Keele (1968), with specific images adapted from a similar category learning task developed by Fried and Holyoak (1984) (and later used in an imaging study by Seger et al., 2000). Participants were asked to classify abstract visual stimuli into one of two categories. Participants had to guess initially but received feedback after each decision. The stimuli consisted of exemplars (10 by 10 grids of squares in two colors; Fig. 1d) from two categories; each category was defined by a prototype stimulus. Each exemplar differed from its prototype in the color of 7 tiles that were randomly selected from the 100-tile grid. Within each prototype, half the tiles were one color, and half the other color. Additionally, each pair of prototypes shared 50 tiles (50% of color-location combinations) with each other Each of the two versions were composed of 2 distinct categories generated with the same principles using distinct prototypes. Version A prototypes were taken directly from previous studies (Fried & Holyoak, 1984; Seger et al., 2000); we developed Version B prototypes using MatLab following the exact parameters of version A. The prototype stimuli and exemplar stimuli for each condition, can be seen in the Supplemental materials (Supplemental Figs. 2 and 3). For each version, participants classified 50 exemplars from each category for a total of 100 trials. The order of exemplar presentation was random but constant across participants. The dependent measure was the percent of correctly classified trials. All implicit learning tasks were implemented in PsychoPy (Peirce, 2009).

2.3.2. Measures of explicit memory and general intelligence

2.3.2.1. Kaufman brief intelligence test. The Kaufman Brief Intelligence Test provides measures of verbal, non-verbal, and composite IQ (Kaufman & Kaufman, 2004). The test itself consists of a multiple-choice picture vocabulary test, a matrix completion test, and an open-ended verbal riddles test. Visual stimuli for the picture and matrix tests are included in a test booklet provided by the publisher of the test. We used the Verbal IQ, Non-Verbal IQ, and Composite IQ measures calculated from these subtests as predictors of individual performance on other tasks.

2.3.2.2. California Verbal Learning Test-Version II. Following the standard protocol, participants were read a list of 16 words and then asked immediately to recall as many words as possible; this procedure—reading a list of words and asking for immediate recall—was repeated five times with the same word list. The sum of correct responses from these five trials was used as the dependent outcome. (The full California Verbal Learning Test provides additional types of measures, but we used only the immediate recall sum). The Standard Version of the California Verbal Learning Test-II was used as Form A and the Alternate Version as Form B; these differ only in the content of the word lists (they are identical in format, procedure, etc.). Both versions were created and published by Woods, Delis, Scott, Kramer, and Holdnack (2006). This instrument was chosen in part for its previously established high reliability.

2.3.2.3. Reading span. To assess working memory, we used a Reading Span Test (Conway et al., 2005; Oswald, McAbee, Redick, & Hambrick, 2015; Redick et al., 2012) and administered via Millisecond Software/Inquisit. In this test, participants must hold 3–7 items (letters) in memory while assessing the truth-validity of a subsequently presented sentence on each trial. Under the absolute scoring method, trials in which the participant correctly recalls all the memory items (letters) receive a score of 1, and all other trials receive a score of 0. The sum of these scores is the absolute "storage score." The span can further be described as the proportion correct out of possible correct, giving a score between 0 and 1 based on the sum of trials scored as above.

2.4. Data analysis

2.4.1. Task performance, version and session effects

All analyses were conducted using Stata 14 (revision 29, Statacorp, 2018). Differences from chance were invested using one-sample t-tests. To test for version and session effects, we conducted linear regression analyses separately for each task using the learning scores as the outcome and the following categorical variables as predictors: session (Session 1 or Session 2), version (A or B), and a session-by-version interaction. Analyzing the data this way is equivalent to performing fully crossed 2 (Version) × 2 (Sessions) ANOVA, but with the advantage of providing effect size estimates in the form of standardized betas, as well as estimates of differences across levels of a factor without having to perform post-hoc tests. See the Supplemental Materials for exact model specifications for each model.

2.4.2. Test-retest reliability and cross-task relationships

Prior to performing correlations across tasks and test sessions, we applied a standard linear equating method to all version B scores. First, the Version B scores were z-transformed. Then, using the mean and standard deviation of version A scores, these version B z-scores were transformed into the same scale of A. This method of linear equating is commonly used in psychometrics to equate scores across tests of different difficulties (Kolen & Brennan, 2004). By linear equating, we essentially removed the effect of version from the Session 1 – Session 2 comparison. See the Supplemental Material for exact formulas.

To assess test-retest reliability, we calculated the Pearson product-moment correlation between the scores after linear equating. We

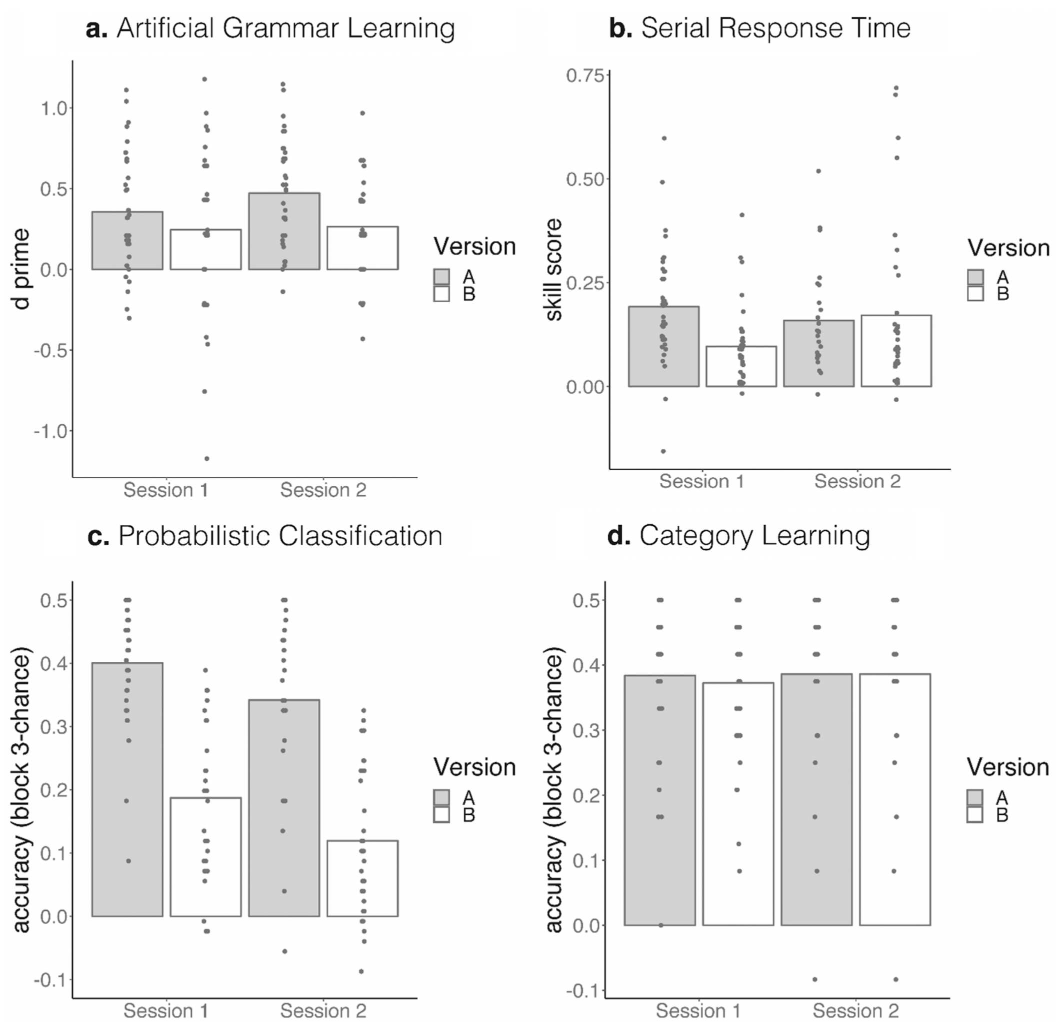

Fig. 2. Learning measures for each implicit learning task by version and session. (a) Discriminability (d prime) is plotted for artificial grammar learning. In this and all other graphs each point represents an individual participant's score and bars depict the group mean for the session and version. (b) Skill scores ($\frac{MeanRT_{RandomBlocks(3,7)} - MeanRT_{SequenceBlocks(5,6)}}{MeanRT_{RandomBlocks(3,7)}}$) are plotted for the serial reaction time task. (c) Block 3 percent correct-50% scores are plotted for the probabilistic classification task. (d) Block 3 percent correct-50% scores are plotted for the implicit category learning task. Greater values for Skill and Learning scores indicate more learning.

examined cross-task correlations using Pearson correlation for the T1 scores only, after linear equating. The magnitudes of certain Pearson correlations were compared with Fischer's r-to-z transformations.

2.4.3. Factor analyses

For each of the factor analysis models, we used the standardized (z-transformed) Session 1 scores only for each included task and constrained the model to retain one factor. A Bayesian information criterion (BIC) was used to compare models (the model with the lowest BIC is preferred) (Schwarz, 1978). See the Supplemental Material for exact model specifications.

2.4.4. Modeling learning strategies during the probabilistic classification task

Because the probabilistic classification task can be performed using an array of different learning strategies (Gluck, Shohamy, & Myers, 2002) some of which are thought to be more explicit, we classified participants by their strategy use on this task. To classify participants, we used a method of fitting models based on trial-by-trial data (available here: http://www.meeter.nl/m/). This method provides the best-fitting model for each participant overall, as well as the best-fitting model for each participant within successive 24-trial windows. Because we used the participants' performance in the third block of trials to determine learning scores, we used the modal best-fitting model for the windows corresponding to the trials in the third block to classify participants by strategy. This method classifies participants as having as many as 14 different strategies, based on the cues that are associated most strongly with their trial by trial responses. These 14 strategies can be collapsed into 3 strategy categories identified by Gluck et al. (2002): One-cue (responses based on one cue combination only); Singleton (responses based on the cue combinations consisting of only one cue); and Multi-cue. While it is theoretically possible for any of these strategies to be executed using either explicit or implicit learning mechanisms, the one-cue and singleton strategies are suggestive of overt memorization and explicit learning, while the multi-cue strategies are difficult to execute explicitly (Gluck et al., 2002; Meeter, Myers, Shohamy, Hopkins, & Gluck, 2006; Meeter, Radicsa, Myers, Gluck, & Hopkins, 2008). The classification method may also fail to classify a participant's strategy, in which case the participant is classified as "random," although he or she may have simply used a strategy not well fit by any of the included models.

3. Results

Separate analyses which included delay time between testing sessions were conducted for each of the tasks. Since there was no effect of delay time between testing sessions on any of the tests (all $p > .05$, Supplemental Table 10), this factor is not included in subsequent analyses. We did not observe that any participant's performance on any of the 6 measures was more than 3 standard deviations from the mean overall performance during either test session; no outliers were therefore removed.

3.1. Performance on implicit learning tasks.

3.1.1. Artificial Grammar Learning

Across all sessions and conditions, mean discriminability (d-prime) scores were positive ($M = 0.43$, $SD = 0.21$) and significantly different from 0 (one sample t-test: $t\ (77) = 15.97$, $p < .001$), indicating that after training, participants were able to discriminate grammatical from non-grammatical test items. Before comparing performance within subjects, we checked for session (practice) and version effects. To this end, we performed a linear regression with participants' discriminability (d-prime) scores as the outcome (dependent measure) and the following categorical variables as predictors: session (Session 1 or Session 2), version (A or B), and a session-by-version interaction. Discriminability did not differ by session ($\beta_{session} = 0.163$, $t = 1.30$, $p = .19$) or version ($\beta_{version} = 0.03$, $t = 0.28$, $p = .77$), and there was no evidence of a session-by-version interaction ($\beta_{sessionXversion} = -0.058$, $t = -0.38$, $p = .70$) (in this and all subsequent regressions, standardized coefficients are reported). The median d-prime values for each combination of version and session are displayed in Fig. 2a.

3.1.2. Serial reaction time task

Across all sessions and conditions, mean skill scores were positive ($M = 0.18$, $SD = 0.13$ ms) and significantly different from 0 (one sample t-test: $t(64) = 9.60$, $p < .001$), indicating that after training, participants were faster on sequence blocks than adjacent random blocks. In order to check for practice and version effects, we performed a linear regression with participants' skill scores as the outcome and session and version as categorical predictors. There was no main effect of session ($\beta_{session} = -0.5$, $t = 1.22$, $p = .225$). However, skill scores for version A were higher than for version B ($\beta_{version} = -0.11$, $t = 3.25$, $p < .001$), suggesting that version B may have been more difficult for participants. We also observed a session-by-version interaction. In this case, those who trained on version B first improved more during the second session (($\beta_{sessionXversion} = 0.13$, $t = 2.63$, $p = .01$) than those who trained on version A first. Mean skill scores for each session and version are displayed in Fig. 2b.

3.1.2.1. Explicit knowledge of sequence. Twenty-four percent of the sample did not report any awareness of a pattern. An additional 14% responded that they did sense a pattern, but were unable to produce any correctly sequenced items. Across participants who did correctly recall parts of the sequence, the mean number of correct items was 2.5 ($SD = 1.6$) and the range was from 0 to 12 (Supplemental Fig. 8). Previous work using this measure has defined explicit sequence knowledge as having almost complete sequence recall (11 or 12 items out of 12 correct) (Song, Howard, & Howard, 2007). Similarly, Willingham and Goedert-Eschmann (1999) found that participants with no real sequence knowledge were able to guess an average of 4.6 correct items, whereas an explicit (intentional learning) group averaged 8.5 items. Based on this previous work, we defined "high sequence knowledge" as correctly recalling 8 or more items. Eight participants correctly recalled more than 8 items. These participants also had the greatest skill scores at Session 2 and were more than 2 SDs above the mean skill score. As reflected in their skill scores, these participants also displayed extremely low RTs in the final sequence block—less than 300 ms, which is more than 2 SDs away from the mean RT for that block across participants—but not in the subsequent random block.

Analyses excluding these 8 participants follow the same patterns as reported above; mean skill scores in Session 2 were positive ($M = 0.17$, $SD = 0.13$) and significantly different from 0 (one sample t-test: $t\ (28) = 6.9$, $p < .0.001$), indicating that after training, participants were faster on sequence blocks than adjacent random blocks. Practice and version effects were also similar. These are detailed in the Supplemental Material.

3.1.3. Probabilistic classification task

To establish that participants learned the probabilistic associations, we first compared their optimal response rate (percent optimal responses) in each block of trials to chance (50%). For all four blocks (notably including the first block) in both versions, the participants' mean optimal response rates were significantly above chance (minimum one-sample $t > 5.0$, $p < .001$).

We performed a linear regression with the difference between the optimal response rate and chance (50%) on each block of trials as the outcome and the following categorical variables as predictors: block number (1, 2, 3 or 4) session (Session 1 or Session 2) and version (A or B). This analysis produced evidence of a main effect of block number such that participants' rates of optimal responding were most different from chance in blocks after Block 1; in Block 1, the rate of optimal response is expected to be lowest because participants are just beginning to learn cue-response associations in block 1. The difference was greatest for block 3 compared to block 1 ($\beta_{Block2} = 0.11$, $t = 9.33$, $p < .001$; $\beta_{Block3} = 0.12$, $t = 10.28$, $p < .001$; $\beta_{Block4} = 0.11$, $t = 9.16$, $p = < 0.001$).

As has been observed previously (Knowlton et al., 1994), in Block 4 we observed lower scores than Block 3 (Block 3 $M = 0.87$; Block 4 $M = 0.86$, $t = 1.27 p = .21$) and a higher standard deviation of scores as compared to Block 3 (Block 4 $SD = 0.13$, Block 3 $SD = 0.12$), suggesting that participants may have become fatigued and inattentive, or attempted to change strategies in the final block. Based on these differences, as well as the finding that the difference from chance was greatest for Block 3, we used mean accuracy in Block 3 to calculate our learning score for this task. Specifically**,** we subtracted chance performance (50%) from the observed mean accuracy in Block 3 (Block 3 – 50%).

We also found evidence of a session effect such that participants provided the optimal response less often the second time they did the task ($\beta_{session} = -0.03$, $t = -2.19$, $p = .028$) and an effect of version ($\beta_{version} = -0.22$, $t = -14.87$, $p < .001$). There was no session-by-version interaction ($\beta_{sessionXversion} = -0.02$, $t = -1.02$, $p = .307$). Importantly, learning (the block number effect) did not interact with session or version, indicating that the magnitude of learning did not differ across sessions. The mean optimal response difference from chance for each combination of version and session are displayed in Fig. 1c. Despite the differences by Version and Session, based on the differences in performance by Block, it is clear that learning did take place in all conditions.

3.1.3.1. Probabilistic classification task strategy classification. Participants were classified as using multi-cue, one-cue, or singleton strategies; those who were not well-fit by one of those models were labeled "random" by the classification methods. One-cue and singleton strategies are suggestive of overt memorization and explicit learning, while the multi-cue strategies are difficult to execute explicitly

At both Session 1 and Session 2, the majority of participants (Session 1: 68%; Session 2: 47%) were best fit by a multi-cue strategy. At Session 2, a greater number of participants' response patterns failed to be fit by the models and were assigned "random" than at Session 1 (Session 1: 10.9% Session 2: 29.6%; two-sample test of proportions = $z = -2.56$, $p = .01$), but the number of singleton and one-cue participants was similar at both sessions (singleton: Session 1: 10.9%, Session 2: 7.8%; one-cue Session 1: 9.4% Session 2: 15.6%). Accordingly, the proportion of participants using a memorization strategy (one cue or singleton) was not significantly different between Sessions 1 and 2 (Session 1: 20% Session 2: 26%, two-sample test of proportions: $z = 0.78$, $p = .44$).

Notably, we re-ran the analyses reported above excluding anyone who could be fit by a memorization strategy. To establish that participants learned the probabilistic associations, we first compared their optimal response rate (percent optimal responses) in each block of trials to chance (50%). For all four blocks (notably including the first block) in both versions, the participants' mean optimal response rates were significantly above chance (minimum one-sample $t > 5.0$, $p < .001$). We found a similar pattern: a trend toward a session effect such that participants provided the optimal response less often the second time they did the task ($\beta_{session} = -0.14$, $t = -1.81$, $p < .07$) and an effect of version ($\beta_{version} = -0.69$, $t = -9.24$, $p < .001$). There was no session-by-version interaction ($\beta_{sessionXversion} = 0.11$, $t = 1.4$, $p = .162$). There were no block-by-session or block-by-version interactions; thus learning (the block number effect) did not interact with session or version, indicating that the magnitude of learning did not differ across sessions.

3.1.4. Implicit category learning

To establish that participants learned the categories, we first compared their mean accuracy in each block of trials to chance (50%). For all four blocks (notably including the first block) in both versions, the participants' mean accuracies were significantly above chance (minimum one-sample $t > 12.8$, $p < .001$). We performed a linear regression with participants' accuracy on each block of trials as the outcome and the following categorical variables as predictors: block number (1, 2, 3 or 4) session (Session 1 or Session 2) and version (A or B). This analysis produced evidence of a block number effect such that participants were more accurate on blocks 2,3, and 4 than on block 1, and the difference was greatest for block 3 compared to block 1 ($\beta_{Block2} = 0.29$, $t = 5.65$, $p < .001$; $\beta_{Block3} = 0.39$, $t = 7.63$, $p < .001$; $\beta_{Block4} = 0.38$, $t = 7.43$, $p = < 0.001$). We observed lower scores and a greater standard deviation of scores in Block 4 as compared with Block 3 (Block 3 $M = 0.758$, $SD = 0.171$; Block 4 $M = 0.755$, $SD = 0.168$), suggesting that participants may have become fatigued and inattentive in the final block. Based on these differences, as well as the finding that the difference from Block 1 was greatest for Block 3, we used mean accuracy in Block 3 to calculate our learning score for this task. Specifically, we subtracted chance performance (50%) from the observed mean accuracy in Block 3 (Block 3 – 50%). We found no evidence of a session effect ($\beta_{session} = -0.05$, $t = -0.82$, $p = .41$), or a version effect ($\beta_{version} = -0.08$, $t = -1.31$, $p = .19$), or a session-by-version interaction ($\beta_{sessionXversion} = -0.05$, $t = -0.75$, $p = .45$). We also did not observe a significant block-by-session or block-by-version interaction. The mean accuracy difference from chance in Block 3 for each combination of version and session are displayed in Fig. 2d.

3.2. Performance on explicit learning, IQ and working memory tasks.

3.2.1. California Verbal Learning Test

We performed a linear regression on participants' 5-trial sum scores from the California Verbal Learning Test with session and version as predictors, including a test for session-by-version interaction. Session 2 mean scores were higher than session 1 scores ($\beta_{session} = 0.36$, $t = 2.88$, $p = .005$); version B mean scores were higher than version A scores ($\beta_{version} = 0.33$, $t = 2.64$, $p = .009$); however, because of a negative interaction between session and version ($\beta_{sessionXversion} = -0.35$, $t = -2.20$, $p = .03$), version B at session 2 was not the highest of the four mean scores.

Table 1
Fisher's r-to-z comparisons of test re-test reliability (r) between California Verbal Learning Test and implicit learning tasks.

	Test -Retest r	n	Fischer's z (vs. CVLT)	p
CVLT	0.76	62	NA	NA
AGL	0.07	59	4.1	< 0.001
SRT	0.38	52	2.89	0.004
PCT	0.39	62	2.97	0.003
Category	0.44	55	2.57	0.01

3.2.2. KBIT (administered once)

Average IQ measures were high for both subscales and composite IQ: Verbal (M = 113.98, SD = 14.81), Non-Verbal (M = 109.37, SD = 12.66), and Full Scale Standard IQ (M = 110.93, SD = 12.69).

3.2.3. Reading span (administered once)

Participants scored an average of 42.6 ($SD = 18.19$) on the reading span task. For comparison, the normative sample used by Engle and colleagues (Redick et al., 2012) had a mean reading span storage score of 36.51 ($SD = 18.83$).

3.3. Test-retest reliability

In Table 1 we display the Session 1 – Session 2 correlations for learning on the implicit learning tasks (after linear equating for version effects). For artificial grammar learning, we used d-prime as a high d-prime is only possible if participants learned something about the artificial structure/grammar; for the serial reaction time task, we used skill scores; for probabilistic classification and implicit category learning, we used the mean accuracy in Block 3 minus chance (50%). The scatterplots (Fig. 3) for Session 1–-Session 2 correlations did not suggest non-linear relationships or problematic heteroscedasticity, so we used a linear model for correlation.

For serial reaction time task, probabilistic classification, and implicit category learning, the learning measures within individuals were correlated across testing sessions (serial reaction time task: $r = 0.38$, $p = .006$; probabilistic classification task: $r = 0.39$, $p = .002$; category learning: $r = 0.45$, $p < .001$; Fig. 3). However, for artificial grammar learning, learning was not correlated across testing sessions ($r = .07 p = .53$). For California Verbal Learning Test, the Session 1 – Session 2 raw scores (i.e. not t-scaled) were correlated ($r = 0.75$, $p < .001$). Fisher's r-to-z transformation revealed that the difference in the magnitude of correlation between California Verbal Learning Test and each of the implicit learning tasks was significant at the 0.01 alpha level or greater (Table 1). Thus, three of the four implicit learning measures showed test-re-test consistency while artificial grammar learning did not. The magnitude of this consistency was, however, smaller than the magnitude of the test-retest consistency of the explicit measure, the California Verbal Learning Test ($r = 0.76$).

We ran these same analyses excluding participants who were deemed to have explicit knowledge in the serial reaction time task (n = 8) or used an explicit learning strategy in the probabilistic classification task (n = 28). As shown in Supplemental Table 21, for serial reaction time task and probabilistic classification, the learning measures across individuals were still correlated across testing sessions (serial reaction time task: $r = .45$, $p = .005$; probabilistic classification task: $r = 0.38.$, $p = .02$).

3.4. Cross-task relationships

As shown in Table 2, pairwise correlations between implicit learning tasks from low to moderate. Neither working memory nor IQ measures correlated with measures of implicit learning (Table 2).

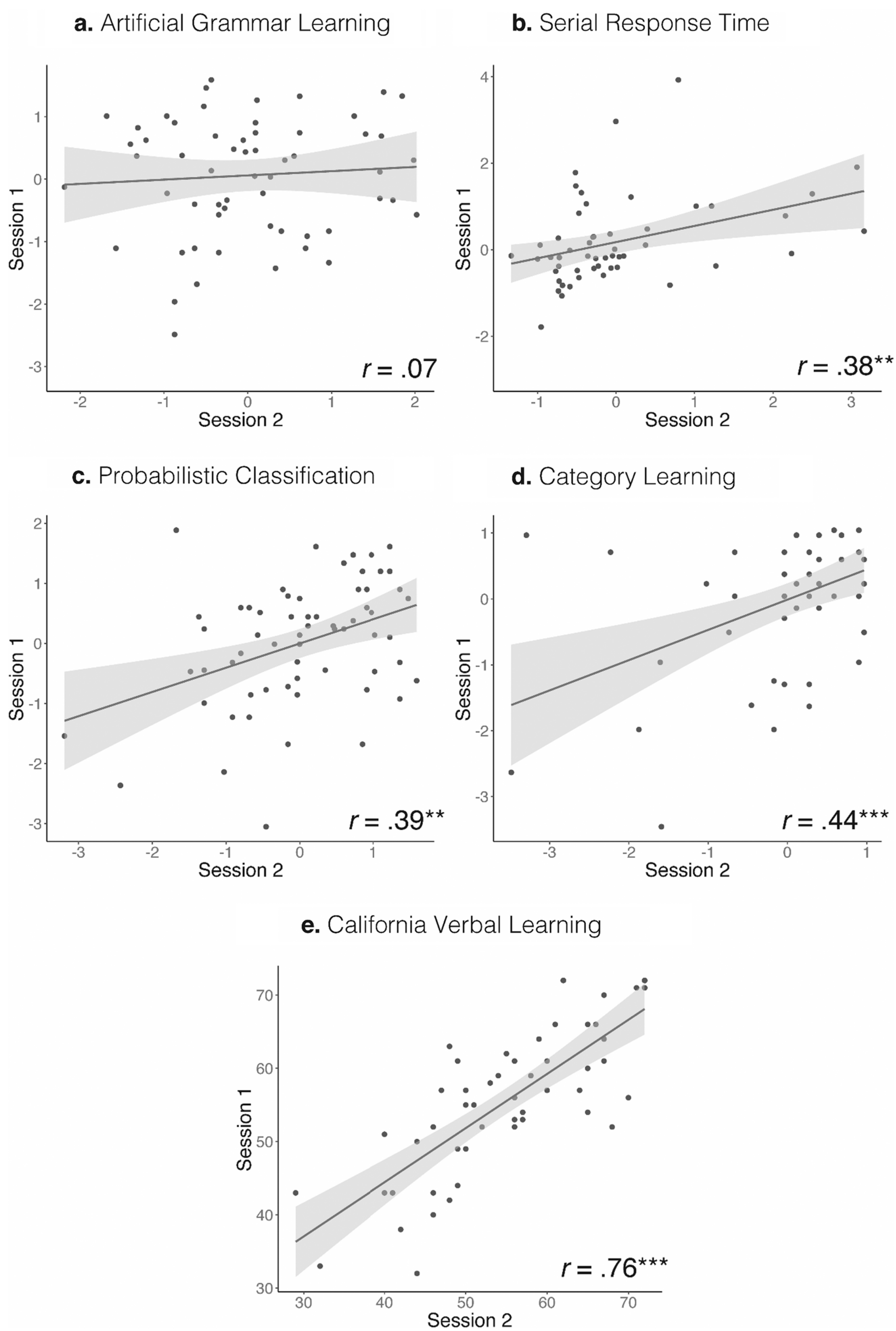

Fig. 3. Implicit and explicit learning measures across test sessions. (a) artificial grammar learning, (b) serial reaction time task, (c) probabilistic classification task, and (d) implicit category learning. (e) Explicit learning across test sessions Each point represents an individual; lines depict the ordinary least squares (OLS) regression; shaded areas denote one standard error of the mean. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

3.5. Factor analysis

We tested a one-factor model using standardized Session 1 scores for artificial grammar learning, probabilistic classification, serial reaction time, and implicit category learning as indicators. The model likelihood ratio fit statistic ($\chi^2(3) = 20.51$, $p = .002$) suggested that this model was a poor fit. Next, we fit a single-factor model using only probabilistic classification, serial reaction time, and implicit category learning as indicators. In contrast to the previous model, this model's likelihood ratio fit statistic (($\chi^2(3) = 3.87$, $p = .276$) suggested a good fit for the

Table 2
Pairwise correlations between working memory, IQ, and implicit learning.

Cross-Task Correlations									
	WM	Verbal IQ	Non-V IQ	Comp. IQ	CVLT	AGL	SRT	PCT	ICL
WM	1.00								
Verbal IQ	0.15	1.00							
Non-V IQ	0.12	0.42*	1.00						
Comp. IQ	0.33*	0.87***	0.81***	1.00					
CVLT	−0.01	0.39**	0.04	0.28	1.00				
Artificial Grammar Learning	−0.16	0.07	−0.11	−0.08	0.20	1.00			
Serial Response Task	0.09	0.10	0.06	0.09	−0.18	−0.17	1.00		
Probabilistic Classification Task	−0.08	0.04	−0.01	0.04	0.28*	0.32**	0.07	1.00	
Category Learning	−0.07	−0.17	0.01	−0.01	0.14	−0.17	0.01	0.24*	1.00

* $p < .05$.
** $p < .01$.
*** $p < .001$.

data. Furthermore, the Bayesian Information Criterion (BIC) for the model that included artificial grammar learning was 25.35, but without artificial grammar learning the model BIC was only 12.43 (lower BIC indicates better fit). The eigenvalues and (unrotated) factor loadings for these models appear in Supplemental Tables 11 and 12 respectively.

We also attempted to fit a one-factor model with IQ, WM, and implicit learning measures, but no one-factor model accommodated all these variables well; for example, a model including serial reaction time, probabilistic classification, and implicit category learning with WM and IQ had a likelihood ratio statistic of $\chi^2(10) = 14.69$, $p = .14$ and BIC of 21.54 compared with the model including only the three implicit learning measures which had a likelihood ratio statistic of $\chi^2(10) = 3.87$, $p = .276$ and BIC of 12.43 (Supplemental Tables 13 and 14).

4. Discussion

We discovered that reliable individual differences in implicit learning do exist for several different measures of implicit learning (probabilistic classification, category learning, and sequence learning). Probabilistic classification, implicit category learning, and sequence learning all showed medium (by Cohen's guidelines) Session 1 to Session 2 correlations within participants. However, the Session 1 – Session 2 correlation (and by inference the reliability of individual differences) in artificial grammar learning was low and not statistically significant. Further, although the three implicit learning tasks differed considerably among themselves in cognitive and perceptual-motor demands, learning on the three tasks appeared related because a single-factor model supported the possibility of a shared underlying factor in learning. Finally, implicit learning was not associated with ability in working memory or IQ.

The explicit learning task (California Verbal Learning Test) showed higher test-retest reliability than any of the implicit learning tasks, but this finding was not unexpected for three reasons. First, like any published psychometric instrument, the California Verbal Learning Test has been extensively tested and tuned before publication for the specific purpose of increasing reliability; this is not the case for experimental tasks in general. Second, in some of our tasks (such as the serial reaction time task), the measurement of learning is based on a difference score (for example between random and sequence block RT); difference scores may inherently have lower reliability (Cronbach, 1970; Lord, 1958) (although some disagree—see Rogosa & Willett, 1983; Thomas & Zumbo, 2012). Third, the stability of working memory capacity and explicit learning ability have been well established in the literature (e.g., Rönnlund et al., 2005; Waters & Caplan, 2003; Xu et al., 2018).

While the California Verbal Learning Test showed higher test-re-test reliability (i.e., stability) as compared with the implicit learning tasks, it is not possible to directly compare the variability across individuals on this explicit memory measure to the variability across individuals on implicit measures by comparing the variances. As previously noted (Reber, 1993), there is no statistically valid method for directly comparing the variances of two different latent variables. Nevertheless, low variability across individuals was an important part of Reber's initial hypothesis that implicit learning differed fundamentally from the well-known individual variability in explicit learning. He asserted that implicit learning—being an evolutionarily old learning mechanism—would be expected to function in a relatively narrow range across individuals (Reber, 1993). Because of this, observed individual differences at one time point would be more likely to reflect error rather than real underlying differences in ability. While we cannot directly compare the variability across our implicit and explicit measures, our findings show that implicit learning is stable across individuals over time. Given this, variation across individuals within any given test is unlikely to simply reflect measurement error.

Our findings are also partially consistent with the findings of a previous study on the test-retest reliability of implicit learning (Stark-Inbar et al., 2017). In that study, significant test-retest reliability was found for a visuo-motor adaptation task and an alternating version of the serial reaction time task, but not for the non-alternating serial reaction time task. Because we used a non-alternating serial reaction time task and did find significant test-retest reliability, this point deserves some scrutiny. First, we note that a linear equating method was not used to account for differences in difficulty between the two sequences in the Stark-Inbar et al. (2017) study, which could artificially deflate test-retest reliability. Second, significant test-retest reliability was observed on the non-alternating serial reaction time task when learning was measured mid-task, but not when learning was measured near the end of the task. It is possible that participants became fatigued or inattentive late in learning. Notably, the version of the non-alternating serial reaction time task in the prior study was roughly twice as long as the version in the present study. The Stark-Inbar et al. (2017), study also tested both versions during the same day, so fatigue may have been more likely. Taken together, the present results are largely consistent with this previous work.

Moreover, our work extends the observations of reliable individual differences to additional indices of implicit learning that are outside of the perceptual motor domain. In particular, probabilistic classification and implicit category learning are generally accepted to be examples of cognitive skill learning while serial reaction time tasks could be driven by perceptual, spatial, or motor mechanisms (Dennis, Howard, & Howard, 2006; Goschke & Bolte, 2012; Willingham, Wells, Farrell, & Stemwedel, 2000). Despite differences in the type of learning, we found that learning on probabilistic classification, implicit category, and serial reaction time tasks all loaded on a common factor and showed similar patterns of stable individual differences. This finding supports the idea of a domain-general mechanism for implicit learning.

Although we found levels of learning and performance similar to those reported in previous studies for artificial grammar learning (e.g. Reber & Squire, 1999; Reber et al., 2003; Seger et al., 2000), this task—perhaps the (former) archetype of implicit learning paradigms—did not display reliable individual differences and did not load on a factor with the other implicit learning measures. One possible explanation stems from differences in task demands between artificial grammar learning and the other implicit learning tasks: the artificial grammar learning paradigm requires explicit study in the learning phase and a direct response in the test phase (i.e., participants are asked to judge "is this grammatical or not"). This explicit aspect may "contaminate" artificial grammar learning as a measure of implicit learning at the second time-point. We know from many years of research with the artificial grammar learning that it is of course not completely explicit; if performing the artificial grammar learning consisted of some part implicit learning (with some degree of reliability) and some part explicit learning (with its own degree of reliability), then the test-retest reliability of artificial grammar learning as a whole should be positive. However, it is possible that re-testing on this paradigm could produce a shift in strategy and the balance of implicit and explicit processes that are deployed for learning. It is possible that individuals performed the artificial grammar learning task differently on each occasion, contributing to the lack of test-retest reliability we observed.

Another interpretation is that the underlying mechanism serving artificial grammar learning is in fact quite different from that underlying the serial reaction time, probabilistic classification, and implicit category learning tasks. This interpretation is supported by the neuropsychology literature. Serial reaction time task, probabilistic classification, and implicit category learning (as well as artificial grammar learning) are often *preserved* in amnesia and Alzheimer's disease (Gabrieli, 1998; Heindel et al., 2013; Knowlton & Squire, 1993; Knowlton et al., 1992, 1994; Knowlton et al., 1996; Nosofsky et al., 2012; Reber et al., 2003)—meaning that medial temporal-lobe structures are not required for learning on these tasks. Moreover, serial reaction time task, probabilistic classification, and implicit category learning are often *impaired* in diseases of the basal ganglia, such as Parkinson's and Huntington's diseases (Ashby et al., 2005; Foerde & Shohamy, 2011a,2011b; Knowlton et al., 1996; Knowlton et al., 1996; Shohamy et al., 2004; Smith & McDowall, 2006). In contrast, artificial grammar learning appears to be *intact* in disorders of the basal ganglia (Knowlton et al., 1996; Smith & McDowall, 2006; Smith, Siegert, & McDowall, 2001; Witt, Nuhsman, & Deuschl, 2002), suggesting that it may be executed by a striatal-independent system or a flexible combination/hybrid of systems. Crucially, Reber's prediction of a general lack of stable individual differences is borne out in the case of artificial grammar learning but not the other implicit learning tasks; this may not be coincidental given that Reber used artificial grammar learning almost exclusively in his studies of implicit learning.

Although it is not possible to guarantee that the other measures of implicit learning are entirely process-pure (e.g., Gluck et al., 2002, Willingham & Goedert-Eschmann, 1999), we examined the possibility of explicit process contributions to our observed implicit learning stability in several ways. Performance on the serial reaction time, category learning and probabilistic classification tasks did not correlate with performance on working memory or IQ, making it improbable that Session 1-Session 2 correlations were driven by these explicit abilities. However, to further confirm that the reliability was not being driven by explicit learning on the tasks, we examined the relationship between implicit and explicit learning in the serial reaction time task and classified the most likely strategies for participants in the probabilistic classification task. For the serial reaction time task, we found that explicit sequence learning was very low overall. In addition, for the probabilistic classification task, the majority of participants were best fit by a multi-cue strategy at both time points, which supports the interpretation that probabilistic classification scores reflect implicit learning. We also re-ran all analyses on the serial reaction time data and the procedural classification data excluding participants who were deemed to have explicit knowledge, and all of the same cross-task correlations and test-re-test results were obtained despite reductions in the overall power. We did not measure explicit knowledge on implicit category learning, but participants did not report any overt patterns in post-test interviews, nor did any participants deviate from mean performance to the same extent as the high explicit knowledge serial reaction time task subjects. Indeed, the high explicit-knowledge serial reaction time task participants had skill scores greater than two standard deviations above the mean; no category learning participants deviated from the mean this much. Furthermore, these tests have been shown to be independent of explicit knowledge in previous work (Eldridge, Masterman, & Knowlton, 2002; Knowlton et al., 1996; Reber, Knowlton, & Squire, 1996; Squire & Knowlton, 1995).

Our conclusion that reliable implicit learning reflects a stable trait, of course, merits further investigation. Future work should expand upon measures of this construct with an even wider array of tasks and over longer periods of time. We have shown reliability over the course of a week—extending a previous investigation that looked during the same day—but it would be useful to determine how stable these individual differences are over longer periods of time.

The current findings raise an interesting issue for theories of intelligence that reduce it to a simple global parameter, such as processing speed or efficiency (e.g. Karmiloff-Smith & Thomas, 2003). This is a corollary of the finding that implicit learning is not related to IQ and working memory but nevertheless exhibits stable individual differences. If both IQ and implicit learning were driven by a global parameter, then we would expect them to be correlated and to load on a common factor for that parameter. However, implicit measures did not correlate with IQ or working memory (Table 2) and we did not find any feasible model in which IQ and any measure of implicit learning loaded on a common factor (see Supplemental Materials, Tables 13 and 14). These data therefore provide evidence for the existence of a completely uncorrelated cognitive ability, negating the "global" aspect of the postulated parameter (e.g. why would processing speed affect IQ but not implicit learning?).

To conclude, we now know that implicit learning ability across a range of tasks (but not all tasks that have been considered measures of implicit learning) can be reliably measured across time in individuals. Because the observed differences in implicit learning reflect real differences in ability, we now have the potential to apply this knowledge to training paradigms for a variety of tasks in which implicit learning is implicated, from second-language learning to medical diagnosis (Aldridge, Glodzik, Ballerini, Fisher, & Rees, 2011; Ullman, 2001). Whether domain-general implicit learning mechanisms can be trained, and the manner in which explicit learning can help or hinder implicit learning and subsequent performance remain open questions (Finn et al., 2014; Sanchez & Reber, 2013)—questions that now have a foundation to build on. Moreover, despite the age- and IQ-independence of implicit learning, the suggestion that implicit learning mechanisms could be used as a compensatory mechanism (see e.g. Ullman & Pullman, 2015) for students who struggle with explicit learning (such as students with low IQ or WM spans) must now take into account the role of individual differences in implicit learning.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.05.007.

References

Aldridge, R. B., Glodzik, D., Ballerini, L., Fisher, R. B., & Rees, J. L. (2011). Utility of non-rule-based visual matching as a strategy to allow novices to achieve skin lesion diagnosis. *Acta Dermato-Venereologica, 91*, 279–283.

Amso, D., & Davidow, J. (2012). The development of implicit learning from infancy to

adulthood: Item frequencies, relations, and cognitive flexibility. *Developmental Psychobiology, 54*, 664–673.
Ashby, F. G., Noble, S., Filoteo, J., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology, 17*, 115–124. https://doi.org/10.1037/0894-4105.17.1.115.
Ashby, G., Valentin, V., Johnson, S., & Stout, J. (2005). Category learning in Huntington's disease. *Journal of Cognitive Neuroscience, S*, 237.
Atwell, J. A., Conners, F. A., & Merrill, E. C. (2003). Implicit and explicit learning in young adults with mental retardation. *American Journal on Mental Retardation, 108*, 56–68. https://doi.org/10.1352/0895-8017(2003) 108 < 0056:IAELIY > 2.0.CO;2.
Barnes, K. A., Howard, J. H., Jr., Howard, D. V., Kenealy, L., & Vaidya, C. J. (2010). Two forms of implicit learning in childhood ADHD. *Developmental Neuropsychology, 35*, 494–505. https://doi.org/10.1080/875656412010494750.
Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83*, 62–78.
Bhakuni, R., & Mutha, P. K. (2015). Learning of bimanual motor sequences in normal aging. *Frontiers in Aging Neuroscience, 7*. https://doi.org/10.3389/fnagi.2015.00076.
Brown, R. M., Robertson, E. M., & Press, D. Z. (2009). Sequence skill acquisition and off-line learning in normal aging. *PLoS ONE, 4*. https://doi.org/10.1371/journal.pone.0006683.
Bussy, G., Charrin, E., Brun, A., Curie, A., & des Portes, V. (2011). Implicit procedural learning in fragile X and down syndrome. *Journal of Intellectual Disability Research, 55*, 521–528. https://doi.org/10.1111/j.1365-2788.2011.01410.x.
Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Oxford, England: Houghton Mifflin.
Churchill, J. D., Stanis, J. J., Press, C., Kushelev, M., & Greenough, W. T. (2003). Is procedural memory relatively spared from age effects? *Neurobiology of Aging, 24*, 883–892. https://doi.org/10.1016/S0197-4580(02)00194-X.
Cohen, N. J., & Poldrack, R. A. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory, 5*, 131–178.
Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review, 12*(5), 769–786.
Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119.
Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). Oxford, England: Harper.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957.
Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research-Psychologische Forschung, 60*, 24–41. https://doi.org/10.1007/BF00419678.
Dennis, N. A., Howard, J. H., & Howard, D. V. (2006). Implicit sequence learning without motor sequencing in young and old adults. *Experimental Brain Research, 175*, 153–164. https://doi.org/10.1007/s00221-006-0534-3.
Drag, L. L., & Bieliauskas, L. A. (2010). Contemporary review 2009: Cognitive aging. *Journal of Geriatric Psychiatry and Neurology, 23*, 75–93. https://doi.org/10.1177/0891988709358590.
Eldridge, L. L., Masterman, D., & Knowlton, B. J. (2002). Intact implicit habit learning in Alzheimer's disease. *Behavioral Neuroscience, 116*, 722–726. https://doi.org/10.1037//0735-7044.116.4.722.
Ell, S. W., Weinstein, A., & Ivry, R. B. (2010). Rule-based categorization deficits in focal basal ganglia lesion and Parkinson's disease patients. *Neuropsychologia, 48*, 2974–2986. https://doi.org/10.1016/j.neuropsychologia.2010.06.006.
Fernandez-Ruiz, J., Hall, C., Vergara, P., & Diaz, R. (2000). Prism adaptation in normal aging: Slower adaptation rate and larger aftereffect. *Cognitive Brain Research, 9*, 223–226. https://doi.org/10.1016/S0926-6410(99)00057-9.
Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. E. (2015). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology, 142*, 212–220. https://doi.org/10.1016/j.jecp.2015.09.027.
Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When it hurts (and helps) to try: The role of effort in language learning. *PLoS ONE, 9*, e101806.
Fletcher, J., Maybery, M. T., & Bennett, S. (2000). Implicit learning differences: A question of developmental level? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 246–252. https://doi.org/10.1037//0278-7393.26.1.246.
Foerde, K., Poldrack, R. A., & Knowlton, B. J. (2007). Secondary-task effects on classification learning. *Memory & Cognition, 35*, 864–874.
Foerde, K., & Shohamy, D. (2011a). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience, 31*, 13157–13167.
Foerde, K., & Shohamy, D. (2011b). The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. *Neurobiology of Learning and Memory, 96*, 624–636.
Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*(2), 234–257.
Gabrieli, J. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87–115. http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.49.1.87.
Gabrieli, J. D. E., Corkin, S., Mickel, S. F., & Growdon, J. (1993). Intact acquisition and long-term retention of mirror-tracing skill in Alzheimer's disease and in global amnesia. *Behavioral Neuroscience, 107*, 899–910.
Galea, J. M., Albert, N. B., Ditye, T., & Miall, R. C. (2010). Disruption of the dorsolateral prefrontal cortex facilitates the consolidation of procedural skills. *Journal of Cognitive Neuroscience, 22*(6), 1158–1164. https://doi.org/10.1162/jocn.2009.21259.
Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates implicit and explicit learning. *Journal of Experimental Psychology-Learning Memory and Cognition, 33*, 34–54.
Ghetti, S., Angelini, L., & Annunzio, G. D. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development, 79*, 339–358.
Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task? Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*, 408–418. https://doi.org/10.1101/lm.45202.
Goschke, T., & Bolte, A. (2012). On the modularity of implicit sequence learning: Independent acquisition of spatial, symbolic, and manual sequences. *Cognitive Psychology, 65*, 284–320. https://doi.org/10.1016/j.cogpsych.2012.04.002.
Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*, 501–518. http://www.ncbi.nlm.nih.gov/pubmed/3160813.
Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Verissimo, J., Dye, C. D., ... Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities, 32*, 2362–2375. https://doi.org/10.1016/j.ridd.2011.07.026.
Heindel, W. C., Festa, E. K., Ott, B. R., Landy, K. M., & Salmon, D. P. (2013). Prototype learning and dissociable categorization systems in Alzheimer's disease. *Neuropsychologia, 51*, 1699–1708. https://doi.org/10.1016/j.neuropsychologia.2013.06.001.
Horan, W. P., Green, M. F., Knowlton, B. J., Wynn, J. K., Mintz, J., & Nuechterlein, K. H. (2008). Impaired implicit learning in schizophrenia. *Neuropsychology, 22*, 606–617. https://doi.org/10.1037/a0012602.
Howard, J. H., Jr., & Howard, D. V. (2013). Aging mind and brain: Is implicit learning spared in healthy aging? *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00817.
Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science, 15*, 496–505.
Karmiloff-Smith, A., & Thomas, M. (2003). Connectionist models of development, developmental disorders, and individual differences. In R. J. Sternberg, J. Lautrey, & T. I. Lubart (Eds.). *Models of intelligence: International perspectives*. American Psychological Association.
Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321–340. https://doi.org/10.1016/j.cognition.2010.05.011.
Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test: KBIT 2.* Pearson.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996a). A neostriatal habit learning system in humans. *Science, 273*, 1399–1402. http://www.ncbi.nlm.nih.gov/pubmed/8703077.
Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science, 3*, 172–179.
Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262*, 1747–1749. http://www.ncbi.nlm.nih.gov/pubmed/8259522.
Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*, 106–120. https://doi.org/10.1101/lm.1.2.106.
Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996b). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology, 10*, 538–548.
Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.
Kuerten, J., De Vries, M. H., Kowal, K., Zwitserlood, P., & Floeel, A. (2012). Age affects chunk-based, but not rule-based learning in artificial grammar acquisition. *Neurobiology of Aging, 33*, 1311–1317.
Laasonen, M., Vare, J., Oksanen-Hennah, H., Leppamaki, S., Tani, P., Harno, H., ... Cleeremans, A. (2014). Project DyAdd: Implicit learning in adult dyslexia and ADHD. *Annals of Dyslexia, 64*, 1–33. https://doi.org/10.1007/s11881-013-0083-y.
Lin, Y. (2004). Research on mechanisms and individual differences of implicit learning: A developmental perspective. *International Journal of Psychology, 39*, 560.
Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement, 18*(3) Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/001316445801800301.
Lukács, A., & Kemény, F. (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology, 28*, 472–483. https://doi.org/10.1037/neu0000052.
Lum, J. A. G., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities, 34*, 3460–3476. https://doi.org/10.1016/j.ridd.2013.07.017.
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
Maddox, W. T., Aparicio, P., Marchant, N. L., & Ivry, R. B. (2005). Rule-based category learning is impaired in patients with Parkinson's disease but not in patients with cerebellar disorders. *Journal of Cognitive Neuroscience, 17*, 707–723. https://doi.org/10.1162/0898929053747630.
Meeter, M., Myers, C. E., Shohamy, D., Hopkins, R. O., & Gluck, M. A. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning & Memory, 13*(2), 230–239. https://doi.org/10.1101/lm.43006.
Meeter, M., Radicsa, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience and Biobehavioral Reviews, 32*(2), 237–248. https://doi.org/10.1016/j.neubiorev.

2007.11.001.
Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., & Vicari, S. (2008). Structural correlates of implicit learning deficits in subjects with developmental dyslexia. In G. F. Eden, & D. L. Flower (Vol. Eds.), *Learning, skill acquisition, reading, and dyslexia: Vol. 1145*, (pp. 212–221). . http://doi.org/10.1196/annals.1416.010.
Meulemans, T., & Van der Linden, M. (2000). Artificial grammar learning in amnesia. *Consciousness and Cognition, 9*, S88.
Meulemans, T., & Van der Linden, M. (2003). Implicit learning of complex information in amnesia. *Brain and Cognition, 52*, 250–257.
Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology, 69*, 199–221.
Midford, R., & Kirsner, K. (2005). Implicit and explicit learning in aged and young adults. *Aging Neuropsychology and Cognition, 12*, 359–387. https://doi.org/10.1080/13825580500246894.
Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*.
Nosofsky, R. M., Denton, S. E., Zaki, S. R., Murphy-Knudsen, A. F., & Unverzagt, F. W. (2012). Studies of implicit prototype extraction in patients with mild cognitive impairment and early Alzheimer's disease. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38*, 860–880. https://doi.org/10.1037/a0028064.
Ofen, N., Kao, Y.-C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S., & Gabrieli, J. D. E. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience, 10*, 1198–1205. https://doi.org/10.1038/nn1950.
Oswald, F. O., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*.
Pavlidou, E. V., Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of Dyslexia, 59*, 55–77.
Peirce (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. https://doi.org/10.3389/neuro.11.010.2008.
Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550. https://doi.org/10.1038/35107080.
Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews, 32*, 197–205. https://doi.org/10.1016/j.neubiorev.2007.07.007.
Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*, 564–574. https://doi.org/10.1037//0894-4105.13.4.564.
Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005). The neural correlates of motor skill automaticity. *The Journal of Neuroscience, 25*, 5356–5364. https://doi.org/10.1523/JNEUROSCI.3880-04.2005.
Posner, M., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3), https://doi.org/10.1037/h0025953.
Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*(2), 227–244. https://doi.org/10.1037/0033-2909.133.2.227.
Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences, 11*, 274–279.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219–235. https://doi.org/10.1037//0096-3445.118.3.219.
Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford University Press.
Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia, 51*, 2026–2042. https://doi.org/10.1016/j.neuropsychologia.2013.06.019.
Reber, P. J., Knowlton, B. J., & Squire, L. R. (1996). Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience, 110*, 861–871. https://doi.org/10.1037//0735-7044.110.5.861.
Reber, P. J., Martinez, L. A., & Weintraub, S. (2003). Artificial grammar learning in Alzheimer's disease. *Cognitive Affective & Behavioral Neuroscience, 3*, 145–153.
Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson's disease. *Behavioral Neuroscience, 113*, 235–242. https://doi.org/10.1037//0735-7044.113.2.235.
Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888–896.
Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123.
Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology Learning Memory and Cognition, 20*(3), 585–594.
Rieckmann, A., & Backman, L. (2009). Implicit learning in aging: Extant patterns and new directions. *Neuropsychology Review, 19*, 490–503.
Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability the difference score in the measurement of change. *Journal of Educational Measurement, 20*(4), 335–343. https://doi.org/10.1111/j.1745-3984.1983.tb00211.x.
Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging, 20*(1), 3–18.
Rosas, R., Ceric, F., Tenorio, M., Mourgues, C., Thibaut, C., Hurtado, E., & Aravena, M. T. (2010). ADHD children outperform normal children in an artificial grammar implicit learning task: ERP and RT evidence. *Consciousness and Cognition, 19*, 341–351.
Rüsseler, J., Gerth, I., & Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology, 28*, 808–827. https://doi.org/10.1080/13803390591001007.
Sanchez, D. J., & Reber, P. J. (2013). Explicit pre-training instruction does not improve implicit perceptual-motor sequence learning. *Cognition, 126*, 341–351. https://doi.org/10.1016/j.cognition.2012.11.006.
Schugens, M. M., Daum, I., Spindler, M., & Birbaumer, N. (1997). Differential effects of aging on explicit and implicit memory. *Aging Neuropsychology and Cognition, 4*, 33–44.
Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H., & Gabrieli, J. D. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia, 38*(9), 1316–1324.
Seger, C., Prabhakaran, V., Poldrack, R. A., & Gabrieli, J. D. E. (2000). Neural activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. *Psychobiology, 28*, 283–292.
Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*, 676–686. https://doi.org/10.1037/0735-7044.118.4.676.
Smith, J. G., & McDowall, J. (2006). When artificial grammar acquisition in Parkinson's disease is impaired: The case of learning via trial-by-trial feedback. *Brain Research, 1067*, 216–228. https://doi.org/10.1016/j.brainres.200S.10.025.
Smith, J., Siegert, R. J., & McDowall, J. (2001). Preserved implicit learning on both the serial reaction time task and artificial grammar in patients with Parkinson's disease. *Brain and Cognition, 45*, 378–391.
Song, S., Howard, J. H., & Howard, D. V. (2007). Implicit probabilistic sequence learning is independent of explicit awareness. *Learning & Memory, 14*, 167–176. https://doi.org/10.1101/lm.437407.
Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences, 92*, 12470–12474. https://doi.org/10.1073/pnas.92.26.12470.
Staels, E., & Van den Broeck, W. (2017). A specific implicit sequence learning deficit as an underlying cause of dyslexia? Investigating the role of attention in implicit learning tasks. *Neuropsychology, 31*, 371–382. https://doi.org/10.1037/neu0000348.
Stark-Inbar, A., Raza, M., Taylor, J. A., & Ivry, R. B. (2017). Individual differences in implicit motor learning: Task specificity in sensorimotor adaptation and sequence learning. *Journal of Neurophysiology, 117*, 412–428. https://doi.org/10.1152/jn.01141.2015.
Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience, 16*, 1339–1351. https://doi.org/10.1162/0898929042304688.
Thomas, K. M., & Nelson, C. A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology, 79*, 364–387. https://doi.org/10.1006/jecp.2000.2613.
Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA. *Educational and Psychological Measurement, 72*(1), 37–43. https://doi.org/10.1177/0013164411409929.
Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*(10), 717–726.
Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews, 51*, 205–222. https://doi.org/10.1016/J.NEUBIOREV.2015.01.008.
Vakil, E., & Agmon-Ashkenazi, D. (1997). Baseline performance and learning rate of procedural and declarative memory tasks: Younger versus older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52*(5), P229–P234. https://doi.org/10.1093/geronb/52B.5.P229.
Vinter, A., & Detable, C. (2008). Implicit and explicit motor learning in children with and without Down's syndrome. *British Journal of Developmental Psychology, 26*, 507–523. https://doi.org/10.1348/026151007X267300.
Vloet, T. D., Marx, I., Kahraman-Lanzerath, B., Zepf, F. D., Herpertz-Dahlmann, B., & Konrad, K. (2010). Neurocognitive performance in children with ADHD and OCD. *Journal of Abnormal Child Psychology, 38*, 961–969. https://doi.org/10.1007/s10802-010-9422-1.
Voytek, B., & Knight, R. T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proceedings of the National Academy of Sciences of the United States of America, 107*(42), 18167–18172. https://doi.org/10.1073/pnas.1007277107.
Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus, 24*(8), 920–933. https://doi.org/10.1002/hipo.22279.
Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*(4), 550–564.
Willingham, D. B., & Dumas, J. A. (1997). Long-term retention of a motor skill: Implicit sequence knowledge is not retained after a one-year delay. *Psychological Research, 60*(August 1996), 113–119.
Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological Science, 10*, 531–536.
Willingham, D. B., Greeley, T., Bardone, A. M., Baker, D., Cohen, A., & Nissen, M. J. (1993). Dissociation in a serial response time task using a recognition measure: Comment on Perruchet and Amorim (1992). *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 19, 1424–1430.

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology, 88*, 1451–1460. http://www.ncbi.nlm.nih.gov/pubmed/12205165.

Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition, 28*, 366–375. https://doi.org/10.3758/BF03198552.

Witt, K., Nuhsman, A., & Deuschl, G. (2002). Intact artificial grammar learning in patients with cerebellar degeneration and advanced Parkinson's disease. *Neuropsychologia, 40*, 1534–1540.

Woods, S., Delis, D., Scott, J., Kramer, J., & Holdnack, J. (2006). The California Verbal Learning Test – second edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology, 21*(5), 413–420. https://doi.org/10.1016/j.acn.2006.06.002.

Xu, Z., Adam, K. C. S., Fang, X., & Vogel, E. K. (2018). The reliability and stability of visual working memory capacity. *Behavior Research Methods, 50*(2), 576–588.

Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review, 11*(6), 1048–1054. https://doi.org/10.3758/BF03196735.

Zannino, G. D., Perri, R., Zabberoni, S., Caltagirone, C., Marra, C., & Carlesimo, G. A. (2012). Performing prototype distortion tasks requires no contribution from the explicit memory systems: Evidence from amnesic MCI patients in a new experimental paradigm. *Neuropsychologia, 50*(12), 2907–2915.